Zero-Sum Stochastic Stackelberg Games

Anonymous Author(s)
Affiliation
Address
email

Abstract

Min-max optimization problems (i.e., zero-sum games) have been used to model problems in a variety of fields in recent years, from machine learning to economics. The literature to date has mostly focused on static zero-sum games, assuming independent strategy sets. In this paper, we study a form of dynamic zero-sum games, called stochastic games, with dependent strategy sets. Just as zero-sum games with dependent strategy sets can be interpreted as zero-sum Stackelberg games, stochastic zero-sum games with dependent strategy sets can be interpreted as zero-sum stochastic Stackelberg games. We prove the existence of an optimal solution in zero-sum stochastic Stackelberg games (i.e., a recursive Stackelberg equilibrium), provide necessary and sufficient conditions for a solution to be optimal, and show that a recursive Stackelberg equilibrium can be computed in polynomial time via value iteration. Finally, we show that stochastic Stackelberg games can model the problem of pricing and allocating goods across agents and time; more specifically, we propose a stochastic Stackelberg game whose solutions correspond to a recursive competitive equilibrium in a stochastic Fisher market. We close with a series of experiments which confirm our theoretical results and show how value iteration performs in practice.

Min-max optimization has paved the way for recent progress in a variety of fields, from machine learning to economics. These applications require computing solutions to **the min-max operator**, i.e., solving a **constrained min-max optimization problem**, also known as zero-sum games (with independent strategy sets): i.e., $\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}} f(\boldsymbol{x},\boldsymbol{y})$, where the objective function $f : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ is continuous, and the strategy sets $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{Y} \subset \mathbb{R}^m$ are nonempty and compact. When f is convex-concave, the seminal minimax theorem [1, 2] holds, and such a problem can be interpreted as a simultaneous-move zero-sum game between an outer player $\boldsymbol{x}$ and an inner player $\boldsymbol{y}$, where the solutions $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in \mathcal{X}\times\mathcal{Y}$ of the min-max operator correspond to a Nash equilibrium. More generally, one can consider **zero-sum stochastic games** (with independent strategy sets) $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{Y} \subset \mathbb{R}^m$, nonempty and compact, and a state-dependent payoff function $r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$, for all $\boldsymbol{s} \in \mathcal{S}$, where the players seek to optimize their cumulative discounted payoffs, in expectation. When $r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$ is bounded, continuous, and concave-convex in $(\boldsymbol{x},\boldsymbol{y})$, for all $\boldsymbol{s} \in \mathcal{S}$, these games are guaranteed to have a unique solution [3],[1] while the **optimal policies**, i.e., the per-state collection of solutions to the min-max operator can be interpreted as a **recursive Nash equilibrium** of a zero-sum stochastic game, and can be computed in polynomial time by iterative application of the min-max operator [3]. Zero-sum *stochastic* games generalize zero-sum games from a single state to multiple states, and have found even more applications in a variety of fields [4].

Recently, Goktas and Greenwald [5] studied the computation of a **generalized min-max operator** i.e., solving a **constrained min-max game with dependent strategy sets**: i.e., $\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:h(\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} r(\boldsymbol{x},\boldsymbol{y})$ where, in addition to the aforementioned assumptions, h :

[1] Although Shapley's original results concern payoffs which are bilinear in the outer and inner players' actions, they extend directly to payoffs which are convex-concave in the players' actions.

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

$\mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is continuous. These authors have shown that even more problems of interest can be captured by solutions to the generalized min-max operator. However, a minimax theorem is not guaranteed to hold in this setting; as a result, these problems are best interpreted as zero-sum sequential, i.e., Stackelberg [6], games, and their solutions, as Stackelberg equilibria. One can likewise consider **zero-sum stochastic games with dependent strategy sets** $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{Y} \subset \mathbb{R}^m$, nonempty and compact and a state-dependent payoff function $r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$, as well as a state dependent constraint function $\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$, for all $\boldsymbol{s} \in \mathcal{S}$, where the players seek to optimize their cumulative discounted payoffs, in expectation, while satisfying the constraint $\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \boldsymbol{0}$ at each state $\boldsymbol{s} \in \mathcal{S}$. Such a problem is best interpreted as a zero-sum stochastic Stackelberg game, for which the appropriate solution concept is the **recursive Stackelberg equilibrium (recSE)**. Although very little is known about the properties of such problems, they generalize both min-max games with dependent strategy sets and zero-sum stochastic games (with independent strategy sets), have applications to autonomous driving [7], reach-avoid problems in human-robot interaction [8], and robust optimization in stochastic environments [9], and, as we show, also have applications in markets.

In this paper, we prove the existence of recSE in zero-sum stochastic Stackelberg games, provide necessary and sufficient conditions for a solution to be a recSE, and show that a recSE can be computed in (weakly) polynomial time via value iteration. We further show that stochastic Stackelberg games can be used to solve problems of pricing and allocating goods across agents and time. In particular, we introduce **stochastic Fisher markets**, a stochastic generalization of the Fisher market [10], and a special case of Friesen's [11] financial market model, which itself is a stochastic generalization of the Arrow and Debreu model of a competitive economy [12]. We then prove the existence of recursive competitive equilibrium (recCE) [13] in this model, under the assumption that consumers have continuous and homogeneous utility functions, by characterizing the CE of any stochastic Fisher market as the Stackelberg equilibria of the corresponding stochastic Stackelberg game. Finally, we use value iteration to solve various stochastic Fisher markets, highlighting the issues that value iteration can encounter, depending on the smoothness properties of the utilities.

Related Work Algorithms for min-max optimization problems (i.e., zero-sum games) with independent strategy sets have been extensively studied [14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37] (for a summary see, Section G [5]). Goktas and Greenwald studied zero-sum games with dependent strategy sets, proposing polynomial-time nested gradient descent ascent (GDA) and simultaneous GDA algorithms for such problems.

The computation of Stackelberg equilibrium in dynamic Stackelberg games has been studied in several interesting settings, but always with independent strategy sets. [39] study continuous-time stochastic Stackelberg games with continuous action spaces, and prove existence of a solution in their setting. [40] and [41] study discrete-time stochastic Stackelberg games with discrete action and state spaces, and provide algorithms to solve such games. [42] consider a stochastic Stackelberg game-like market model with n leaders and m followers; they prove the existence of a Stackelberg equilibrium in their model, and provide (without theoretical guarantees) algorithms that converge to such an equilibrium in experiments. Dynamic Stackelberg games [43] have been applied to a wide range of problems, including security [40, 41], insurance provision [44, 45], advertising [46], robust agent design [47], allocating goods across time intertemporal pricing [48].

The study of algorithms that compute competitive equilibria in Fisher markets was initiated by [49], who provided a polynomial-time method for solving these markets assuming linear utilities. More recently, there have been efforts to study markets in dynamic settings [50, 51, 5], in which the goal is to either track the changing equilibrium of a changing market, or minimize some regret-like quantity for the market. The models considered in these earlier works differ from ours as they do not have stochastic structure and do not invoke a dynamic solution concept.

1 Preliminaries

Notation We use caligraphic uppercase letters to denote sets (e.g., $\mathcal{X}$); bold lowercase letters to denote vectors (e.g., $\boldsymbol{p}, \boldsymbol{\pi}$); bold uppercase letters to denote matrices and vector-valued random variables (e.g., $\boldsymbol{X}, \boldsymbol{\Gamma}$)—which one should be clear from context; lowercase letters to denote scalar quantities (e.g., x, γ); and uppercase letters to denote scalar-valued random variables (e.g., X, Γ). We denote the ith row vector of a matrix (e.g., $\boldsymbol{X}$) by the corresponding bold lowercase letter with subscript i (e.g., $\boldsymbol{x}_i$). Similarly, we denote the jth entry of a vector (e.g., $\boldsymbol{p}$ or $\boldsymbol{x}_i$) by the corresponding

Roman lowercase letter with subscript j (e.g., p_j or x_{ij}). We denote functions by a letter: e.g., f if the function is scalar valued, and $\boldsymbol{f}$ if the function is vector valued. We denote the vector of ones of size n by $\mathbf{1}_n$. We denote the set of integers $\{1, \dots, n\}$ by $[n]$, the set of natural numbers by $\mathbb{N}$, the set of real numbers by $\mathbb{R}$. We denote the postive and strictly positive elements of a set by a $+$ and $++$ subscript respectively, e.g., $\mathbb{R}_+$ and $\mathbb{R}_{++}$. We denote the orthogonal projection operator onto a set C by Π_C, i.e., $\Pi_C(\boldsymbol{x}) = \arg\min_{\boldsymbol{y} \in C} \|\boldsymbol{x} - \boldsymbol{y}\|^2$. We denote by $\Delta_n = \{\boldsymbol{x} \in \mathbb{R}^n_+ \mid \sum_{i=1}^n x_i = 1\}$, and by $\Delta(A)$ the set of probability measures on the set A.

A **stochastic Stackelberg game** $(\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mu^{(0)}, r_{\boldsymbol{x}}, r_{\boldsymbol{y}}, \boldsymbol{g}, p, \gamma)$ is a dynamic two-player sequential, i.e., Stackelberg, game where one player we call the outer-player (resp. inner-player) moves through a set of states $\mathcal{S}$ picking an action to play from their continuous set of actions $\mathcal{X} \subset \mathbb{R}^n$ (resp. $\mathcal{Y} \subset \mathbb{R}^m$). Players start the game at an initial state determined by an initial state distribution $\mu^{(0)} : \mathcal{S} \to [0, 1]$ s.t. for all states $\boldsymbol{s} \in \mathcal{S}$, $\mu^{(0)}(\boldsymbol{s}) \geq 0$ denotes the probability of the game being initialized at state $\boldsymbol{s}$. At each state $\boldsymbol{s} \in \mathcal{S}$ the action $\boldsymbol{x} \in \mathcal{X}$ chosen by the outer player determines the set of **feasible** actions $\{\boldsymbol{y} \in \mathcal{Y} \mid \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \mathbf{0}\}$ that the inner player can in turn play. Once the outer and inner players make their moves, they receive payoffs $r_{\boldsymbol{x}} : \boldsymbol{s} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and $r_{\boldsymbol{y}} : \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, respectively, and the game either ends with probability $1 - \gamma$, where $\gamma \in (0, 1)$ is called the **discount factor**, or transitions to a new state $\boldsymbol{s}' \in \mathcal{S}$, according to a **transition** probability function $p : \mathcal{S} \times \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \to [0, 1]$ s.t. $p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \in [0, 1]$ denotes the probability of transitioning to state $\boldsymbol{s}' \in \mathcal{S}$ from state $\boldsymbol{s} \in \mathcal{S}$ when a strategy profile $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{Y}$ is chosen by the players.

In this paper, we focus on stochastic Stackelberg **zero-sum** games $\mathcal{G}^{(0)} = (\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mu^{(0)}, r, \boldsymbol{g}, p, \gamma)$, in which the outer player's loss is the inner player's gain, i.e., $r_{\boldsymbol{x}} = -r_{\boldsymbol{y}}$. We say that a zero-sum stochastic Stackelberg game is **convex-concave** if, for all $\boldsymbol{s} \in \mathcal{S}$, $r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), g_1(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \dots, g_d(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ are convex-concave in $(\boldsymbol{x}, \boldsymbol{y})$. A zero-sum stochastic Stackelberg game reduces to zero-sum stochastic game [3] if for all state-action tuples $(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \in \mathcal{S} \times \mathcal{X} \times \mathcal{Y}$, $\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq 0$. As we will show, it suffices to focus on deterministic policies in which a **policy** for the outer (resp. inner) player is a mapping from states to actions $\boldsymbol{\pi_x} : \mathcal{S} \to \mathcal{X}$ (resp. $\boldsymbol{\pi_y} : \mathcal{S} \to \mathcal{Y}$). The **outcome** of a zero-sum stochastic Stackelberg game $\mathcal{G}^{(0)}$ is a policy profile $(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \in \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}}$ consisting of policies for the outer and inner players, respectively. An outcome $(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \in \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}}$ is said to be **feasible** if, for all states $\boldsymbol{s} \in \mathcal{S}$, $\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{\pi_x}(\boldsymbol{s}), \boldsymbol{\pi_y}(\boldsymbol{s})) \geq 0$. For simplicity, we introduce a function $\boldsymbol{G} : \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}} \to \mathbb{R}^{|\mathcal{S}| \times d}$ such that $\boldsymbol{G}(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) = (\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{\pi_x}(\boldsymbol{s}), \boldsymbol{\pi_y}(\boldsymbol{s})))_{\boldsymbol{s} \in \mathcal{S}}$, and define feasible outcomes as those $(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \in \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}}$ s.t. $\boldsymbol{G}(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \geq \mathbf{0}$. From now on, we assume:

Assumption 1.1. *1. For all states $\boldsymbol{s} \in \mathcal{S}$, the functions $r(\boldsymbol{s}, \cdot, \cdot), g_1(\boldsymbol{s}, \cdot, \cdot), \dots, g_d(\boldsymbol{s}, \cdot, \cdot)$ are continuous in $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{Y}$ with payoffs r bounded, i.e., $\|r\|_\infty \leq \alpha < \infty$, for some $\alpha \in \mathbb{R}_+$, 2. $\mathcal{X}, \mathcal{Y}$ are non-empty and compact, and for all $\boldsymbol{s} \in \mathcal{S}$ and $\boldsymbol{x} \in \mathcal{X}$ there exists $\boldsymbol{y} \in \mathcal{Y}$ s.t. $\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \mathbf{0}$.*[2]

Given a zero-sum stochastic Stackelberg game $\mathcal{G}^{(0)}$, the **state-value function**, $v : \mathcal{S} \times \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}} \to \mathbb{R}$, and the **action-value function**, $q : \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}} \to \mathbb{R}$ are respectively defined as:
$v(\boldsymbol{s}; \boldsymbol{\pi_x}, \boldsymbol{\pi_y}) = \mathbb{E}^{\boldsymbol{\pi_x}, \boldsymbol{\pi_y}}_{\boldsymbol{S}^{(t+1)} \sim p(\cdot \mid \boldsymbol{S}^{(t)}, \boldsymbol{X}^{(t)}, \boldsymbol{Y}^{(t)})} \left[\sum_{t=0}^{\infty} \gamma^t r(\boldsymbol{S}^{(t)}, \boldsymbol{X}^{(t)}, \boldsymbol{Y}^{(t)}) \mid \boldsymbol{S}^{(0)} = \boldsymbol{s}\right]$ and
$q(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\pi_x}, \boldsymbol{\pi_y}) = \mathbb{E}^{\boldsymbol{\pi_x}, \boldsymbol{\pi_y}}_{\boldsymbol{S}^{(t+1)} \sim p(\cdot \mid \boldsymbol{S}^{(t)}, \boldsymbol{X}^{(t)}, \boldsymbol{Y}^{(t)})} \left[\sum_{t=0}^{\infty} \gamma^t r(\boldsymbol{S}^{(t)}, \boldsymbol{X}^{(t)}, \boldsymbol{Y}^{(t)}) \mid \boldsymbol{S}^{(0)} = \boldsymbol{s}, \ \boldsymbol{X}^{(0)} = \boldsymbol{x}, \boldsymbol{Y}^{(0)} = \boldsymbol{y}\right]$.

For clarity, we write expectations conditional on $\boldsymbol{X}^{(t)} = \boldsymbol{\pi_x}(\boldsymbol{S}^{(t)})$ and $\boldsymbol{Y}^{(t)} = \boldsymbol{\pi_y}(\boldsymbol{S}^{(t)})$ as $\mathbb{E}^{\boldsymbol{\pi_x}, \boldsymbol{\pi_y}}$, and denote the state- and action-value functions by $v^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}(\boldsymbol{s})$, and $q^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$, respectively. Additionally, we let $\mathcal{V} = [-\alpha/1-\gamma, \alpha/1-\gamma]^{\mathcal{S}}$ be the space of all state-value functions of the form $v : \mathcal{S} \to [-\alpha/1-\gamma, \alpha/1-\gamma]$, and we let $\mathcal{Q} = [-\alpha/1-\gamma, \alpha/1-\gamma]^{\mathcal{S} \times \mathcal{X} \times \mathcal{Y}}$ be the space of all action-value functions of the form $q : \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \to [-\alpha/1-\gamma, \alpha/1-\gamma]$. Note that by Assumption 1.1 the range of the state- and action-value functions is $[-\alpha/1-\gamma, \alpha/1-\gamma]$. The cumulative payoff function of the game $u : \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}} \to \mathbb{R}$ is the total expected loss (resp. gain) of the outer (resp. inner) player is then given by $u(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) = \mathbb{E}_{\boldsymbol{s} \sim \mu^{(0)}(\boldsymbol{s})} \left[v^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}(\boldsymbol{s})\right]$.

Definition 1.2. *A feasible policy profile $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}}) \in \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}}$ is said to be a **recursive Stackelberg equilibrium (recSE)** of a zero-sum stochastic Stackelberg game $\mathcal{G}^{(0)}$ iff*
$\max_{\boldsymbol{\pi_y} \in \mathcal{Y}^{\mathcal{S}} : \boldsymbol{G}(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi_y}) \geq \mathbf{0}} u\left(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi_y}\right) \leq u\left(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}}\right) \leq \min_{\boldsymbol{\pi_x} \in \mathcal{X}^{\mathcal{S}}} \max_{\boldsymbol{\pi_y} \in \mathcal{Y}^{\mathcal{S}} : \boldsymbol{G}(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \geq \mathbf{0}} u\left(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}\right)$.

Remark 1.3. *This is a strong definition of a recSE, since we require the constraints $\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \geq \mathbf{0}$ to be satisfied at all states $\boldsymbol{s} \in \mathcal{S}$, not only states which are reached with strictly positive probability.*

[2] Note that this condition is weaker than Slater's condition; it simply ensures the feasible action sets are non-empty for the inner player at each state.

Mathematical Preliminaries A probability measure $q_1 \in \Delta(\mathcal{S})$ **convex stochastically dominates (CSD)** $q_2 \in \Delta(\mathcal{S})$ if $\int_{\mathcal{S}} v(s)q_1(s)ds \geq \int_{\mathcal{S}} v(s)q_2(s)ds$ for all continuous, bounded, and convex functions v on S. A transition function p is termed **CSD convex** in $\boldsymbol{x}$ if for all $\lambda \in (0,1)$, $\boldsymbol{y} \in \mathcal{Y}$ and any $(\boldsymbol{s}', \boldsymbol{x}'), (\boldsymbol{s}^\dagger, \boldsymbol{x}^\dagger) \in \mathcal{S} \times \mathcal{X}$, with $(\boldsymbol{s}, \boldsymbol{x}) = \lambda(\boldsymbol{s}', \boldsymbol{x}') + (1-\lambda)(\boldsymbol{s}^\dagger, \boldsymbol{x}^\dagger)$, we have $\lambda p(\cdot \mid \boldsymbol{s}', \boldsymbol{x}', \boldsymbol{y}) + (1-\lambda)p(\cdot \mid \boldsymbol{s}^\dagger, \boldsymbol{x}^\dagger, \boldsymbol{y})$ CSD $p(\cdot \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$. A transition function p is termed **CSD concave** in $\boldsymbol{y}$ if for all $\lambda \in (0,1)$ and any $(\boldsymbol{s}', \boldsymbol{y}'), (\boldsymbol{s}^\dagger, \boldsymbol{y}^\dagger) \in \mathcal{S} \times \mathcal{X} \times \mathcal{Y}$, with $(\boldsymbol{s}, \boldsymbol{y}) = \lambda(\boldsymbol{s}', \boldsymbol{y}') + (1-\lambda)(\boldsymbol{s}^\dagger, \boldsymbol{y}^\dagger)$, we have $p(\cdot \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ CSD $\lambda p(\cdot \mid \boldsymbol{s}', \boldsymbol{x}, \boldsymbol{y}') + (1-\lambda)p(\cdot \mid \boldsymbol{s}^\dagger, \boldsymbol{x}, \boldsymbol{y}^\dagger)$. A mapping $L : \mathcal{A} \to \mathcal{B}$ is said to be a **contraction mapping** (resp. **non-expansion**) w.r.t. norm $\|\cdot\|$ iff for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{A}$, and for $k \in [0,1)$ (resp. $k = 1$) such that $\|L(\boldsymbol{x}) - L(\boldsymbol{y})\| \leq k \|\boldsymbol{x} - \boldsymbol{y}\|$.

2 Properties of Recursive Stackelberg equilibrium

We first state a necessary condition that the state-value function induced by a policy profile needs to satisfy in order for that policy profile to be a recSE. It is not directly clear if there exists a policy for any zero-sum stochastic Stackelberg game which satisfies this necessary condition since there is no way to guarantee that such an inequality will hold at all states.[3]

Lemma 2.1. *Consider a zero-sum stochastic Stackelberg game $\mathcal{G}^{(0)}$. A policy profile $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}}) \in \mathcal{S}^{\mathcal{X}} \times \mathcal{S}^{\mathcal{Y}}$ is a recSE if, for all $\boldsymbol{s} \in \mathcal{S}$, we have that: $\max_{\boldsymbol{y} \in \mathcal{Y}: \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{y}) \geq \boldsymbol{0}} q^{\boldsymbol{\pi}^*_{\boldsymbol{x}} \boldsymbol{\pi}^*_{\boldsymbol{y}}}(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{y}) \leq q^{\boldsymbol{\pi}^*_{\boldsymbol{x}} \boldsymbol{\pi}^*_{\boldsymbol{y}}}(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{x}), \boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{y})) \leq \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}: \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \boldsymbol{0}} q^{\boldsymbol{\pi}^*_{\boldsymbol{x}} \boldsymbol{\pi}^*_{\boldsymbol{y}}}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$. Equivalently, a policy profile $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}})$ is a recSE if $(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}))$ is a Stackelberg equilibrium: i.e., a solution to $\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}: \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \boldsymbol{0}} q^{\boldsymbol{\pi}^*_{\boldsymbol{x}} \boldsymbol{\pi}^*_{\boldsymbol{y}}}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ at each $\boldsymbol{s} \in \mathcal{S}$.*

We define an operator $C : \mathcal{V} \to \mathcal{V}$ associated with a zero-sum stochastic Stackelberg game $\mathcal{G}^{(0)}$ whose fixed points satisfy the condition given in Lemma 2.1, and hence correspond to the value function associated with a recSE of $\mathcal{G}^{(0)}$. We then show that this operator is a contraction mapping, thereby establishing the existence of such a fixed point. This result generalizes a result first shown by [3] for zero-sum stochastic games, i.e., zero-sum stochastic Stackelberg games in which $\boldsymbol{G}(\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}) \geq \boldsymbol{0}$, for all $(\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}) \in \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}}$. Define $C : \mathcal{V} \to \mathcal{V}$ for a stochastic Stackelberg game $\mathcal{G}^{(0)}$ as the operator $(Cv)(\boldsymbol{s}) = \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}: \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \boldsymbol{0}} \mathbb{E}_{\boldsymbol{S}' \sim p(\cdot \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})} [r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) + \gamma v(\boldsymbol{S}')]$.

Theorem 2.2. *$(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}})$ is a recSE of $\mathcal{G}^{(0)}$ of $v^{\boldsymbol{\pi}_{\boldsymbol{x}} \boldsymbol{\pi}_{\boldsymbol{y}}}$ iff it induces a value function which is a fixed point of C: i.e., $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}})$ is a Stackelberg equilbrium iff, for all $\boldsymbol{s} \in \mathcal{S}$, $\left(Cv^{\boldsymbol{\pi}^*_{\boldsymbol{x}} \boldsymbol{\pi}^*_{\boldsymbol{y}}}\right)(\boldsymbol{s}) = v^{\boldsymbol{\pi}^*_{\boldsymbol{x}} \boldsymbol{\pi}^*_{\boldsymbol{y}}}(\boldsymbol{s})$.*

The following technical lemma is crucial to proving that C is a contraction mapping. It tells us that the generalized min-max operator is non-expansive; in other words, the generalized min-max operator is 1-Lipschitz w.r.t. the sup-norm.

Lemma 2.3. *Suppose that $f, h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, $\boldsymbol{g} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$ are continuous functions, and $\mathcal{X}, \mathcal{Y}$ are compact sets. Then $\left|\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}: \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \geq \boldsymbol{0}} f(\boldsymbol{x}, \boldsymbol{y}) - \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}: \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \geq \boldsymbol{0}} h(\boldsymbol{x}, \boldsymbol{y})\right| \leq \max_{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{Y}} |f(\boldsymbol{x}, \boldsymbol{y}) - h(\boldsymbol{x}, \boldsymbol{y})|$.*

With the above lemma in hand, we can now prove that C is a contraction mapping.

Theorem 2.4. *Consider the operator C associated with a stochastic Stackelberg game $\mathcal{G}^{(0)}$. Under Assumption 1.1, C is a contraction mapping w.r.t. to the sup norm $\|.\|_\infty$ with constant γ.*

Given some initial state-value function $v^{(0)} \in \mathcal{V}$, we define the **value iteration** process as $v^{(t+1)} = Cv^{(t)}$, for all $t \in \mathbb{N}_+$ (Algorithm 1). One way to interpret $v^{(t)}$ is as the function that returns the value $v^{(t)}(\boldsymbol{s})$ of each state $\boldsymbol{s} \in \mathcal{S}$ in the t stage zero-sum Stackelberg game starting at the last stage t and continuing until stage 0, with terminal payoffs given by $v^{(0)}$. The following theorem, which is a consequence of Theorems 2.2 and 2.4, not only states the existence of a recSE but also provides us with a means of computing a recSE via value iteration.

Theorem 2.5. *Consider a zero-sum stochastic Stackelberg game $\mathcal{G}^{(0)}$. Under Assumption 1.1, $\mathcal{G}^{(0)}$ has a unique value function $v^{\boldsymbol{\pi}^*_{\boldsymbol{x}} \boldsymbol{\pi}^*_{\boldsymbol{y}}}$ associated with all recSE $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}})$, which can be computed by iteratively applying C to any initial state-value function $v^{(0)} \in \mathcal{V}$: $\lim_{t \to \infty} v^{(t)} = v^{\boldsymbol{\pi}^*_{\boldsymbol{x}} \boldsymbol{\pi}^*_{\boldsymbol{y}}}$.*

[3] All omitted results and proofs can be found in the appendix.

Remark 2.6. *Unlike Shapley's existence theorem for recursive Nash equilibria in zero-sum stochastic games, the above theorem does not require that the payoff function be convex-concave. The only conditions needed are continuity of the payoffs and constraints, and bounded payoffs. This makes the recSE a potentially useful solution concept, even for non-convex-non-concave stochastic games.*

Note that the proof of Theorem 2.5 shows that there exists a policy profile which satisfies the conditions of Lemma 2.1. Since this recSE definition is independent of the initial state distribution, we can infer that the recSE of any zero-sum Stackelberg game $\mathcal{G}^{(0)} = (\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mu^{(0)}, r, \boldsymbol{g}, p, \gamma)$ is independent of the initial state distribution $\mu^{(0)}$. Hence, in the remainder of the paper, we denote a zero-sum Stackelberg game by $\mathcal{G}$.

It seems that a stronger equilibrium concept for zero-sum stochastic Stackelberg games would be one that requires the strategy profile be a recSE of each subgame, aptly called **subgame perfect Stackelberg equilibrium**. A feasible policy profile $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*) \in \mathcal{S}^{\mathcal{X}} \times \mathcal{S}^{\mathcal{Y}}$ is a subgame perfect Stackelberg equilibrium of a zero-sum stochastic Stackelberg game $\mathcal{G}$ if, for all $\boldsymbol{s} \in \mathcal{S}, t \in \mathbb{N}, \mu^{(t+1)}(\boldsymbol{s}) = \sum_{\boldsymbol{s}' \in \mathcal{S}} p(\boldsymbol{s} \mid \boldsymbol{s}', \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}'), \boldsymbol{\pi}_{\boldsymbol{y}}^*(\boldsymbol{s}'))\mu^{(t)}(\boldsymbol{s}')$, $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*)$ is a recSE of $(\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mu^{(t)}, r, \boldsymbol{g}, p, \gamma)$. By definition, the set of subgame perfect Stackelberg equilibria of a zero-sum stochastic Stackelberg game is a subset of the set of recSE of the game. Since the set of recSE is independent of the initial state distribution $\mu^{(0)}$, the reverse inclusion also holds. We thus have the following corollary:

Corollary 2.7. *The set of recSE and subgame perfect Stackelberg equilibria coincide in zero-sum stochastic Stackelberg games.*

Theorem 2.5 tells us that value iteration converges to the value function associated with a recSE. Additionally, under Assumption 1.1, recSE is computable in (weakly) polynomial time.[4]

Theorem 2.8. *[Convergence of Value Iteration] Suppose value iteration is run on input $\mathcal{G}$. Let $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*)$ be recSE of $\mathcal{G}$ with value function $v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}$. Under Assumption 1.1, if we initialize $v^{(0)}(\boldsymbol{s}) = 0$, for all $\boldsymbol{s} \in \mathcal{S}$, then for $k \geq \frac{1}{1-\gamma} \log \frac{2\alpha}{\epsilon(1-\gamma)}$, we have $v^{(k)}(\boldsymbol{s}) - v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}) \leq \epsilon$.*

3 Subdifferential Envelope Theorems and Optimality Conditions for Recursive Stackelberg Equilibrium

In this section, we derive optimality conditions for recursive Stackelberg equilibria. In particular, we provide necessary conditions for a policy profile to be a recSE of any zero-sum stochastic Stackelberg game, and we provide sufficient conditions for a policy profile to be a recursive Stackelberg equilibrium of a zero-sum convex-concave stochastic Stackelberg game. Using these results, we prove in the next section that recursive market equilibrium [13] is an instance of recSE in a large class of stochastic markets.

The Benveniste-Scheinkman theorem characterizes the derivative of the optimal value function associated with a recursive optimization problem w.r.t. its parameters, when it is differentiable [52]. Our proofs of the necessary and sufficient optimality conditions rely on a novel subdifferential generalization (Theorem C.3, Appendix C) of this theorem, which applies even when the optimal value function is not differentiable. A consequence of our subdifferential version of the Benveniste-Scheinkman theorem is that we can easily derive the first-order necessary conditions for a policy profile to be a recSE of any stochastic Stackelberg game $\mathcal{G}$ satisfying Assumption 1.1, under standard regularity conditions.

Theorem 3.1. *Consider a zero-sum stochastic Stackelberg game $\mathcal{G}$, where $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n \mid q_1(\boldsymbol{x}) \leq 0, \ldots, q_p(\boldsymbol{x}) \leq 0\}$ and $\mathcal{Y} = \{\boldsymbol{y} \in \mathbb{R}^m \mid r_1(\boldsymbol{y}) \geq 0, \ldots, r_l(\boldsymbol{y}) \geq 0\}$ are convex. Let $\mathcal{L}_{\boldsymbol{s},\boldsymbol{x}}(\boldsymbol{y}, \boldsymbol{\lambda}) = r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) + \gamma\, \mathbb{E}_{\boldsymbol{S}' \sim p(\cdot \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})} [v(\boldsymbol{S}', \boldsymbol{x})] + \sum_{k=1}^{d} \lambda_k g_k(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$. Suppose that Assumption 1.1 holds, and that 1. for all $\boldsymbol{s} \in \mathcal{S}, \boldsymbol{y} \in \mathcal{Y}$, $\max_{\boldsymbol{y} \in \mathcal{Y}} \left\{ r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) + \gamma\, \mathbb{E}_{\boldsymbol{S}' \sim p(\cdot \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})} [v(\boldsymbol{S}', \boldsymbol{x})] \right\}$ is concave in $\boldsymbol{x}$, 2. $\nabla_{\boldsymbol{x}} r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \nabla_{\boldsymbol{x}} g_1(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \ldots, \nabla_{\boldsymbol{x}} g_d(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$,*

[4] This convergence is only weakly polynomial time, since the computation of the generalized min-max operator applied to an arbitrary continuous function is an NP-hard problem; it is at least as hard as non-convex optimization. If, however, one restricts themselves to convex-concave stochastic Stackelberg games, Stackelberg equilibrium is computable in polynomial time, making the problem much more tractable.

$\nabla_{\boldsymbol{y}} r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \nabla_{\boldsymbol{y}} g_1(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \ldots, \nabla_{\boldsymbol{y}} g_d(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ exist, for all $\boldsymbol{s} \in \mathcal{S}, \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y} \in \mathcal{Y}$, 4. $p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ is continuous and differentiable in $(\boldsymbol{x}, \boldsymbol{y})$, and 5. Slater's condition holds, i.e., $\forall \boldsymbol{s} \in \mathcal{S}, \boldsymbol{x} \in \mathcal{X}, \exists \widehat{\boldsymbol{y}} \in \mathcal{Y}$ s.t. $g_k(\boldsymbol{s}, \boldsymbol{x}, \widehat{\boldsymbol{y}}) > 0$, for all $k = 1, \ldots, d$ and $r_j(\widehat{\boldsymbol{y}}) > 0$, for all $j = 1, \ldots, l$, and $\exists \boldsymbol{x} \in \mathbb{R}^n$ s.t. $q_k(\boldsymbol{x}) < 0$ for all $k = 1 \ldots, p$. Then, there exists $\boldsymbol{\mu}^* : \mathcal{S} \to \mathbb{R}^p_+$, $\boldsymbol{\lambda}^* : \mathcal{S} \times \mathcal{X} \to \mathbb{R}^d_+$, and $\boldsymbol{\nu}^* : \mathcal{S} \times \mathcal{X} \to \mathbb{R}^l_+$ s.t. a policy profile $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}}) \in \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}}$ is a recSE of $\mathcal{G}$ only if it satisfies the following conditions, for all $\boldsymbol{s} \in \mathcal{S}$:

$$\nabla_{\boldsymbol{x}} \mathcal{L}_{\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})}(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}), \boldsymbol{\lambda}^*(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))) + \sum_{k=1}^{p} \mu^*_k(\boldsymbol{s}) \nabla_{\boldsymbol{x}} q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) = 0 \tag{1}$$

$$\nabla_{\boldsymbol{y}} \mathcal{L}_{\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})}(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}), \boldsymbol{\lambda}^*(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))) + \sum_{k=1}^{l} \nu^*_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \nabla_{\boldsymbol{x}} r_k(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) = 0 \tag{2}$$

$$\mu^*_k(\boldsymbol{s}) q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) = 0 \qquad q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \leq 0 \quad \forall k \in [p] \tag{3}$$

$$g_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) \geq 0 \qquad \lambda^*_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) g_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) = 0 \quad \forall k \in [d] \tag{4}$$

$$\nu^*_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \nabla_{\boldsymbol{x}} r_k(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) = 0 \qquad r_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \geq 0 \quad \forall k \in [l] \tag{5}$$

In Theorem 3.1, when Assumption 1 is replaced with "for all $\boldsymbol{s} \in \mathcal{S}, \boldsymbol{x} \in \mathcal{X}$, $r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$, $g_1(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \ldots, g_d(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ are concave in $\boldsymbol{y}$," and Assumption 4 is replaced with "$p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ is continuous, CSD concave in $\boldsymbol{y}$, and differentiable in $(\boldsymbol{x}, \boldsymbol{y})$", Equations (74) to (78) become necessary *and sufficient* optimality conditions. For completeness, the reader can find the necessary and sufficient optimality conditions for convex-concave stochastic Stackelberg games under standard regularity conditions in Theorem C.4 (Appendix C). The proof follows exactly as that of Theorem 2.2.

4 Recursive Market Equilibrium

We now introduce an application of zero-sum stochastic Stackelberg games, which generalizes a well known market model, the Fisher market [10], to a dynamic setting in which buyers not only participate in markets across time, but their wealth persists. A **(static) Fisher market** consists of n buyers and m divisible goods [10]. Each buyer $i \in [n]$ is endowed with a budget $b_i \in \mathcal{B}_i \subset \mathbb{R}_+$ and a utility function $u_i : \mathbb{R}^m_+ \times \mathcal{T}_i \to \mathbb{R}$, which is parameterized by a type $t_i \in \mathcal{T}_i$ that defines a preference relation over the consumption space $\mathbb{R}^m_+$. Each good is characterized by a supply $q_j \in \mathcal{Q}_j \subset \mathbb{R}_+$.

An instance of a Fisher market is then a tuple $\mathcal{M} \doteq (n, m, \mathcal{U}, \boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q})$, where $\mathcal{U} = \{u_1, \ldots, u_n\}$ is a set of utility functions, one per buyer, $\boldsymbol{b} \in \mathbb{R}^n_+$ is the vector of buyer budgets, and $\boldsymbol{q} \in \mathbb{R}^m_+$ is the vector of supplies. When clear from context, we simply denote $\mathcal{M}$ by $(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q})$.

A **stochastic Fisher market** is a dynamic market in which each state corresponds to a static Fisher market: i.e., each state $\boldsymbol{s} \in \mathcal{S}$ is characterized by a tuple $(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q})$. In each state, the buyers choose their allocations $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T \in \mathbb{R}^{n \times m}_+$ and the market determines prices, after which the market terminates with probability $(1 - \gamma)$, or it moves into a new state $\boldsymbol{s}'$ with probability $\gamma p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{X})$.[5] A **stochastic Fisher market with savings** is a stochastic Fisher market in which, at each state, each buyer i, in addition to choosing their allocation, can set aside some **savings** $\beta_i \leq b_i$ to spend at some future state, and transitions depend on their savings choices as well: i.e., $p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{X}, \boldsymbol{\beta})$. We denote a stochastic Fisher market (without or with savings) by $\mathcal{F} \doteq (n, m, \mathcal{U}, \mathcal{S}, \boldsymbol{s}^{(0)}, p, \gamma)$.

Given a stochastic Fisher market $\mathcal{F}$ a **recursive competitive equilibrium (recCE)** [13] is a tuple $(\boldsymbol{X}^*, \boldsymbol{\beta}^*, \boldsymbol{p}^*) \in \mathbb{R}^{n \times m \times \mathcal{S}}_+ \times \mathbb{R}^{n \times \mathcal{S}}_+ \times \mathbb{R}^{m \times \mathcal{S}}_+$, which consists of an allocation, savings, and price system s.t. 1) the buyers are expected utility maximizing, constrained by their savings and spending constraints, i.e., for all buyers $i \in [n]$, $(\boldsymbol{x}^*_i, \beta^*_i)$ is an optimal policy that, for all states $\boldsymbol{s} \doteq (\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) \in \mathcal{S}$, solves the Bellman equation $\nu_i(\boldsymbol{s}) =$

$$\max_{(\boldsymbol{x}_i, \beta_i) \in \mathbb{R}^{m+1}_+ : \boldsymbol{x}_i \cdot \boldsymbol{p}^*(\boldsymbol{s}) + \beta_i \leq b_i} \left\{ u_i(\boldsymbol{x}_i, t_i) + \gamma \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}') \sim p(\cdot \mid \boldsymbol{s}, (\boldsymbol{x}_i, \boldsymbol{X}^*_{-i}(\boldsymbol{s})), (\beta_i, \boldsymbol{\beta}^*_{-i}(\boldsymbol{s})))} [\nu_i(\boldsymbol{t}', \boldsymbol{b}' + \beta_i, \boldsymbol{q}')] \right\},$$

[5] Note that, as is standard in the literature, we assume that prices do not determine the next state since market prices are set by a "fictional auctioneer." There is no market participant who determined prices!

where $\boldsymbol{X}^*_{-i}, \boldsymbol{\beta}^*_{-i}$ denote the allocation and saving systems excluding buyer i; and 2) the market clears in each state so that unallocated goods in each state are priced at 0, i.e., for all $j \in [m]$ and $\boldsymbol{s} \in \mathcal{S}$, $p^*_j(\boldsymbol{s}) > 0 \implies \sum_{i\in[n]} x^*_{ij}(\boldsymbol{s}) = 1$ and $p^*_j(\boldsymbol{s}) \geq 0 \implies \sum_{i\in[n]} x^*_{ij}(\boldsymbol{s}) \leq 1$.

The following theorem shows that the recSE of a stochastic Fisher market with savings are in fact recursive competitive equilibria. We note that if one ignores the savings terms, then the recSE are recursive competitive equilibria of the same market without savings.

Theorem 4.1. *A stochastic Fisher market with savings $\mathcal{F}$ in which $\mathcal{U}$ is a vector of continuous and homogeneous utility functions has at least one recCE. Additionally, the recSE $(\boldsymbol{p}^*, \boldsymbol{X}^*, \boldsymbol{\beta}^*)$ that solves the following Bellman equation corresponds to the recCE of $\mathcal{F}$:*

$$v(\boldsymbol{s}) = \min_{\boldsymbol{p}\in\mathbb{R}^m_+} \max_{(\boldsymbol{X},\boldsymbol{\beta})\in\mathbb{R}^{n\times(m+1)}_+ : \boldsymbol{X}\boldsymbol{p}+\boldsymbol{\beta}\leq\boldsymbol{b}} \sum_{j\in[m]} q_j p_j + \sum_{i\in[n]} (b_i - \beta_i)\log(u_i(\boldsymbol{x}_i, t_i)) + \gamma \mathop{\mathbb{E}}_{(\boldsymbol{t}',\boldsymbol{b}',\boldsymbol{q}')\sim p(\cdot|\boldsymbol{s},\boldsymbol{X},\boldsymbol{\beta})} [v(\boldsymbol{t}', \boldsymbol{b}' + \boldsymbol{\beta}, \boldsymbol{q}')] \tag{6}$$

Remark 4.2. *This result cannot be obtained by modifying the Lagrangian formulation, i.e., the simultaneous-move game form, of the Eisenberg-Gale program, because the savings problem is convex-non-concave, and the existence of recursive Nash equilibrium in deterministic policies requires convex-concave payoffs [4].*

5 Experiments

To better understand the iteration complexity of value iteration, and to understand how it performs in a continuous state and action space, we computed the recursive Stackelberg equilibria of various stochastic Fisher markets without and with savings. We created markets with three different classes of utility functions, each of which endowed the state-value function with different smoothness properties.[6] Let $\boldsymbol{\theta}_i \in \mathbb{R}^m$ be a vector of parameters that describes the utility function of buyer $i \in [n]$. We considered the following (standard) utility function classes: 1. linear: $u_i(\boldsymbol{x}_i) = \sum_{j\in[m]} \theta_{ij} x_{ij}$; 2. Cobb-Douglas: $u_i(\boldsymbol{x}_i) = \prod_{j\in[m]} x_{ij}^{\theta_{ij}}$; and 3. Leontief: $u_i(\boldsymbol{x}_i) = \min_{j\in[m]} \left\{\frac{x_{ij}}{\theta_{ij}}\right\}$.

We studied two different experimental setups. We first modeled a stochastic Fisher market *without* savings, where buyers' budgets are replenished at each state. To this end, we assumed that each buyer owned a good sold in the market, with their budget in the next time period proportional to the amount of money spent during the current time period on their good. In this case, we solve the min-max operator using **max-oracle gradient descent (GD)** with an exact max-oracle ([5]; Algorithm 2). Second, we modeled stochastic Fisher markets *with* savings, where buyers budgets, which are initialized at the start of the game, can persist across states, but are never replenished. As these markets are convex-non-concave, our methods are not guaranteed to converge to a recCE. Nonetheless, in attempt to solve them, we used two methods: 1. **nested gradient descent ascent (GDA)** ([5]; Algorithm 3); and 2. **max-oracle gradient descent (GD)** ([5]; Algorithm 4), where we use simulated annealing [53], a heuristic that aims to find a global optimum, as the max-oracle. Since the state space is continuous (budgets, which are continuous, are a part of the state space) we enhanced all algorithms with fitted value iteration, finding a fit via linear regression (e.g., [54]).

In both experimental setups, to check whether the value function found was optimal, we measured the exploitability of the market, meaning the distance between the recCE computed and the actual recCE. To do so, we checked two conditions: 1) whether each buyer's expected utility is maximized at the computed allocation and savings at the prices output by the algorithm, and 2) whether the market always clears. For both settings, given the value function computed by value iteration, we extracted the greedy policy and unrolled it across time to obtain the greedy actions $(\boldsymbol{X}^{(t)}, \boldsymbol{\beta}^{(t)}, \boldsymbol{p}^{(t)})$ at each state $\boldsymbol{s}^{(t)}$. We then computed the cumulative utility of the allocation and savings computed by the algorithms, i.e., for all $i \in [n]$, $\hat{u}_i \doteq \sum_{t=0}^{T} \gamma^t u_i(\boldsymbol{x}_i^{(t)})$. We compared this value to the expected maximum utility u_i^* for each individual buyer i, given the prices and the other buyers' allocations computed by our algorithms. We report the normalized distance between these two values, $\hat{u}_i$ and u_i^*, which we call the **normalized distance to utility maximization (UM)**. For example, in the case of two buyers, we report the normalized distance to UM $= \frac{||(u_1,u_2)-(u_1^\star,u_2^\star)||}{||(u_1^*,u_2^*)||}$. Finally, we

[6] Our code can be found here, and details of our experimental setup can be found in Appendix E.

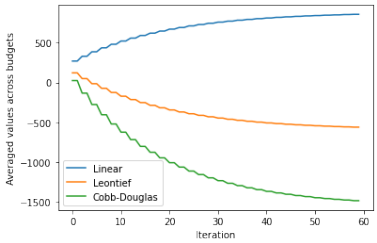

Figure 1: Value function, averaged across all states (i.e., budgets), with fitted value function estimated using 25 randomly sampled states.

Utility Class	Distance to UM	Distance to MC
Linear	9×10^{-9}	2×10^{-8}
Leontief	8×10^{-7}	3×10^{-6}
Cobb-Douglas	1×10^{-2}	1×10^{-4}

Table 1: Exploitability

Table 2: Max-oracle GD in stochastic Fisher markets without savings.

also measured the excess demand, which we took as the **distance to market clearance (MC)**, i.e., $\frac{1}{T}\sum_{t=1}^{T}||\sum_{i\in[n]}\boldsymbol{x}_i^{(t)} - \boldsymbol{q}^{(t)}||$.

In the setup without savings, Figure 1 depicts the average value of the value function across a sample of states as it varies with time, and Table 1 records the exploitability of the recCE computed by our algorithm. For all three class of utility functions, not only do the value functions converge, exploitability is also minimized, as all the buyer utilities are maximized and market always clears.

In the setup with savings (Figure 2), we can see that value iteration converges for both nested GDA and max-oracle GD in Cobb-Douglas markets. Max-oracle GD achieves a higher value, because the oracle does a better job at finding a global maximum for the inner player. In linear markets, only the max-oracle gradient descent algorithm seems to converge. This is most likely due to the non-differentiability of the payoffs in linear Fisher markets, making it difficult for nested GDA to find a global optimum. In Leontief markets, neither algorithm seems to affect the value function very much, which might lead one to conclude that value iteration converges. In terms of exploitability shown in Table 3 and Table 4, however, both algorithms perform more or less as they do in linear markets, which instead suggests that the initial value function is close to optimal.

Surprisingly, even though value iteration converges in both settings for Cobb-Douglas markets, the allocation and savings computed seem to be far from optimal. This is most likely due to the fact that we used linear regression to approximate the value function at each state, even though the value function in Cobb-Douglas markets is strictly concave. In contrast, the buyers' value functions are linear in linear and Leontief markets, and hence compatible with our fitted linear regression.

Utility Class	Normalized Distance to UM	Distance to MC
Linear (Nested GDA)	6.31267	1.51361
Leontief (Nested GDA)	0.40879	2.18851
Cobb-Douglas (Nested GDA)	142.71734	2.18925

Table 3: Exploitability of recCE computed by Nested GDA in Fisher markets with savings.

Utility Class	Normalized Distance to UM	Distance to MC
Linear (Max-oracle GD)	0.38894	2.16509
Leontief (Max-oracle GD)	2.17849	2.16509
Cobb-Douglas (Max-oracle GD)	142.03567	2.16972

Table 4: Exploitability of recCE computed by max-oracle GD in Fisher markets with savings.

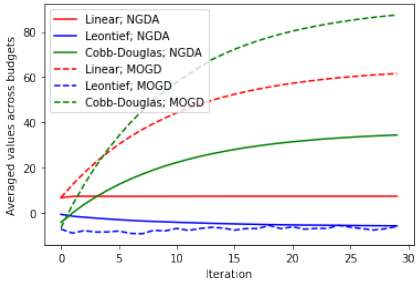

Figure 2: The average value of the value function under Nested GDA (solid line) and Max-oracle GD with simulated annealing (dashed line).

6 Conclusion

In this paper, we proved the existence of recursive Stackelberg equilibria in zero-sum stochastic Stackelberg games, provided necessary and sufficient conditions for a policy profile to be a recursive Stackelberg equilibrium, and showed that a Stackelberg equilibrium can be computed in (weakly) polynomial time via value iteration. Further, we provided a characterization of recursive competitive equilibria in stochastic Fisher markets with and without savings, as they are both instances of zero-sum stochastic Stackelberg games. Finally, we used value iteration to solve for recursive competitive equilibria in these markets. Future work in this space could try using deep reinforcement learning methods to learn better (i.e., nonlinear) representations of the value functions. It is also conceivable that deep reinforcement learning may be able to learn better policies, thereby resolving the difficulties that our methods have in solving for global optima in stochastic Fisher markets with savings.

References

[1] J v Neumann. "Zur theorie der gesellschaftsspiele". In: *Mathematische annalen* 100.1 (1928), pp. 295–320.

[2] Maurice Sion et al. "On general minimax theorems." In: *Pacific Journal of mathematics* 8.1 (1958), pp. 171–176.

[3] Lloyd S Shapley. "Stochastic games". In: *Proceedings of the national academy of sciences* 39.10 (1953), pp. 1095–1100.

[4] Anna Jaśkiewicz and Andrzej S Nowak. "Zero-sum stochastic games". In: *Handbook of dynamic game theory* (2018), pp. 1–64.

[5] Denizalp Goktas and Amy Greenwald. "Convex-Concave Min-Max Stackelberg Games". In: *Advances in Neural Information Processing Systems* 34 (2021).

[6] Heinrich Von Stackelberg. *Marktform und gleichgewicht*. J. springer, 1934.

[7] Jaime F Fisac et al. "Reach-avoid problems with time-varying dynamics, targets and constraints". In: *Proceedings of the 18th international conference on hybrid systems: computation and control*. 2015, pp. 11–20.

[8] Somil Bansal et al. "Hamilton-jacobi reachability: A brief overview and recent advances". In: *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE. 2017, pp. 2242–2253.

[9] Dimitris Bertsimas, David B Brown, and Constantine Caramanis. "Theory and applications of robust optimization". In: *SIAM review* 53.3 (2011), pp. 464–501.

[10] William C Brainard, Herbert E Scarf, et al. *How to compute equilibrium prices in 1891*. Citeseer, 2000.

[11] Peter H. Friesen. "The Arrow-Debreu Model Extended to Financial Markets". In: *Econometrica* 47.3 (1979), pp. 689–707. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1910415.

[12] Kenneth Arrow and Gerard Debreu. "Existence of an equilibrium for a competitive economy". In: *Econometrica: Journal of the Econometric Society* (1954), pp. 265–290.

[13] R Mehra and EC Prescott. "Recursive Competitive Equilibria and Capital Asset Pricing". In: *Essays in Financial Economics* (1977).

[14] Paul Tseng. "On linear convergence of iterative methods for the variational inequality problem". In: *Journal of Computational and Applied Mathematics* 60.1 (1995). Proceedings of the International Meeting on Linear/Nonlinear Iterative Methods and Verification of Solution, pp. 237–252. ISSN: 0377-0427. DOI: https://doi.org/10.1016/0377-0427(94)00094-H. URL: https://www.sciencedirect.com/science/article/pii/037704279400094H.

[15] Yurii Nesterov and Laura Scrimali. "Solving strongly monotone variational and quasi-variational inequalities". In: *Discrete & Continuous Dynamical Systems* 31.4 (2011), pp. 1383–1396.

[16] Gauthier Gidel et al. *A Variational Inequality Perspective on Generative Adversarial Networks*. 2020. arXiv: 1802.10551 [cs.LG].

[17] Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. *Convergence Rate of $\mathcal{O}(1/k)$ for Optimistic Gradient and Extra-gradient Methods in Smooth Convex-Concave Saddle Point Problems*. 2020. arXiv: 1906.01115 [math.OC].

[18] Adam Ibrahim et al. "Lower bounds and conditioning of differentiable games". In: *arXiv preprint arXiv:1906.07300* (2019).

[19] Mingyi Hong, Junyu Zhang, Shuzhong Zhang, et al. *On Lower Iteration Complexity Bounds for the Saddle Point Problems*. 2020. arXiv: 1912.07481 [math.OC].

[20] Tianyi Lin, Chi Jin, and Michael I Jordan. "Near-optimal algorithms for minimax optimization". In: *Conference on Learning Theory*. PMLR. 2020, pp. 2738–2779.

[21] Mohammad Alkousa et al. *Accelerated methods for composite non-bilinear saddle point problem*. 2020. arXiv: 1906.03620 [math.OC].

[22] Anatoli Juditsky, Arkadi Nemirovski, et al. "First order methods for nonsmooth convex large-scale optimization, ii: utilizing problems structure". In: *Optimization for Machine Learning* 30.9 (2011), pp. 149–183.

[23] Erfan Yazdandoost Hamedani and Necdet Serhat Aybat. "A primal-dual algorithm for general convex-concave saddle point problems". In: *arXiv preprint arXiv:1803.01401* 2 (2018).

[24] Zhao Renbo. *Optimal algorithms for stochastic three-composite convex-concave saddle point problems*. 2019. arXiv: 1903.01687 [math.OC].

[25] Kiran Koshy Thekumparampil et al. *Efficient Algorithms for Smooth Minimax Optimization*. 2019. arXiv: 1907.01543 [math.OC].

[26] Yuyuan Ouyang and Yangyang Xu. *Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems*. 2018. arXiv: 1808.02901 [math.OC].

[27] Arkadi Nemirovski. "Prox-method with rate of convergence O (1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems". In: *SIAM Journal on Optimization* 15.1 (2004), pp. 229–251.

[28] Yurii Nesterov. "Dual extrapolation and its applications to solving variational inequalities and related problems". In: *Mathematical Programming* 109.2 (2007), pp. 319–344.

[29] Paul Tseng. "On accelerated proximal gradient methods for convex-concave optimization". In: *submitted to SIAM Journal on Optimization* 1 (2008).

[30] Maziar Sanjabi et al. "On the Convergence and Robustness of Training GANs with Regularized Optimal Transport". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. NIPS'18. Montreal, Canada: Curran Associates Inc., 2018, pp. 7091–7101.

[31] Maher Nouiehed et al. "Solving a class of non-convex min-max games using iterative first order methods". In: *arXiv preprint arXiv:1902.08297* (2019).

[32] Songtao Lu, Ioannis Tsaknakis, and Mingyi Hong. "Block alternating optimization for non-convex min-max problems: algorithms and applications in signal processing and communications". In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 4754–4758.

[33] Chi Jin, Praneeth Netrapalli, and Michael I. Jordan. *What is Local Optimality in Nonconvex-Nonconcave Minimax Optimization?* 2020. arXiv: 1902.00618 [cs.LG].

[34] Dmitrii M Ostrovskii, Andrew Lowy, and Meisam Razaviyayn. "Efficient search of first-order nash equilibria in nonconvex-concave smooth min-max problems". In: *arXiv preprint arXiv:2002.07919* (2020).

[35] Tianyi Lin, Chi Jin, and Michael Jordan. "On gradient descent ascent for nonconvex-concave minimax problems". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 6083–6093.

[36] Renbo Zhao. *A Primal Dual Smoothing Framework for Max-Structured Nonconvex Optimization*. 2020. arXiv: 2003.04375 [math.OC].

[37] Hassan Rafique et al. *Non-Convex Min-Max Optimization: Provable Algorithms and Applications in Machine Learning*. 2019. arXiv: 1810.02060 [math.OC].

[38] Denizalp Goktas and Amy Greenwald. *Robust No-Regret Learning in Min-Max Stackelberg Games*. 2022.

[39] Alain Bensoussan, Shaokuan Chen, and Suresh P Sethi. "The maximum principle for global solutions of stochastic Stackelberg differential games". In: *SIAM Journal on Control and Optimization* 53.4 (2015), pp. 1956–1981.

[40] Deepanshu Vasal. *Stochastic Stackelberg games*. 2020. arXiv: 2005.01997 [math.OC].

[41] Yevgeniy Vorobeychik and Satinder Singh. "Computing stackelberg equilibria in discounted stochastic games". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 26. 1. 2012, pp. 1478–1484.

[42] Victor DeMiguel and Huifu Xu. "A stochastic multiple-leader Stackelberg model: analysis, computation, and application". In: *Operations Research* 57.5 (2009), pp. 1220–1235.

[43] Tao Li and Suresh P Sethi. "A review of dynamic Stackelberg game models". In: *Discrete & Continuous Dynamical Systems-B* 22.1 (2017), p. 125.

[44] Lv Chen and Yang Shen. "On a new paradigm of optimal reinsurance: a stochastic Stackelberg differential game between an insurer and a reinsurer". In: *ASTIN Bulletin: The Journal of the IAA* 48.2 (2018), pp. 905–960.

[45] Yu Yuan, Zhibin Liang, and Xia Han. "Robust reinsurance contract with asymmetric information in a stochastic Stackelberg differential game". In: *Scandinavian Actuarial Journal* (2021), pp. 1–28.

[46] Xiuli He, Ashutosh Prasad, and Suresh P Sethi. "Cooperative advertising and pricing in a dynamic stochastic supply chain: Feedback Stackelberg strategies". In: *PICMET'08-2008 Portland International Conference on Management of Engineering & Technology*. IEEE. 2008, pp. 1634–1649.

[47] Sean C Rismiller, Jonathan Cagan, and Christopher McComb. "Stochastic stackelberg games for agent-driven robust design". In: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. Vol. 84003. American Society of Mechanical Engineers. 2020, V11AT11A039.

[48] Bernt Oksendal, Leif Sandal, and Jan Uboe. "Stochastic Stackelberg equilibria with applications to time-dependent newsvendor models". In: *Journal of Economic Dynamics and Control* 37.7 (2013), pp. 1284–1299.

[49] N. R. Devanur et al. "Market equilibrium via a primal-dual-type algorithm". In: *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.* 2002, pp. 389–395. DOI: 10.1109/SFCS.2002.1181963.

[50] Yun Kuen Cheung, Martin Hoefer, and Paresh Nakhe. "Tracing Equilibrium in Dynamic Markets via Distributed Adaptation". In: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. 2019, pp. 1225–1233.

[51] Yuan Gao, Christian Kroer, and Alex Peysakhovich. "Online Market Equilibrium with Application to Fair Division". In: *Advances in Neural Information Processing Systems* 34 (2021).

[52] L. M. Benveniste and J. A. Scheinkman. "On the Differentiability of the Value Function in Dynamic Models of Economics". In: *Econometrica* 47.3 (1979), pp. 727–732. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1910417.

[53] Dimitris Bertsimas and John Tsitsiklis. "Simulated annealing". In: *Statistical science* 8.1 (1993), pp. 10–15.

[54] Justin Boyan and Andrew Moore. "Generalization in reinforcement learning: Safely approximating the value function". In: *Advances in neural information processing systems* 7 (1994).

[55] Richard Bellman. "On the theory of dynamic programming". In: *Proceedings of the National Academy of Sciences of the United States of America* 38.8 (1952), p. 716.

[56] Stefan Banach. "Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales". In: *Fund. math* 3.1 (1922), pp. 133–181.

[57] Alp E. Atakan. “Stochastic Convexity in Dynamic Programming”. In: *Economic Theory* 22.2 (2003), pp. 447–455. ISSN: 09382259, 14320479. URL: http://www.jstor.org/stable/25055693.

[58] Harley Flanders. “Differentiation Under the Integral Sign”. In: *The American Mathematical Monthly* 80.6 (1973), pp. 615–627. ISSN: 00029890, 19300972. URL: http://www.jstor.org/stable/2319163.

[59] HW Kuhn and AW Tucker. *Proceedings of 2nd berkeley symposium*. 1951.